

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
APPLICATION FOR LETTERS PATENT

Methods and Systems for Processing XML Documents

Inventor(s):

Christopher J. Lovett
Nanshan Zeng

ATTORNEY'S DOCKET NO. MS1-383US

TECHNICAL FIELD

This invention relates to methods and systems for processing Extensible Markup Language (XML) documents. More particularly, the invention concerns ways of validating XML data against a predetermined schema and building, from valid XML data, an in-memory tree representation of the XML document.

BACKGROUND

Extensible Markup Language (XML) is a meta-markup language that provides a format for describing structured data. XML is similar to HTML in that it is a tag-based language. By virtue of its tag-based nature, XML defines a strict tree structure or hierarchy. XML is a derivative of Standard Generalized Markup Language (SGML) that provides a uniform method for describing and exchanging structured data in an open, text-based format. XML utilizes the concepts of elements and namespaces. Compared to HTML, which is a display-oriented markup language, XML is a general-purpose language for representing structured data without including information that describes how to format the data for display.

XML “elements” are structural constructs that consist of a start tag, an end or close tag, and the information or content that is contained between the tags. A “start tag” is formatted as “<tagname>” and an “end tag” is formatted as “</tagname>”. In an XML document, start and end tags can be nested within other start and end tags. All elements that occur within a particular element must have their start and end tags occur before the end tag of that particular element. This defines a tree-like structure that is representative of the XML document. Each element forms a node in this tree, and potentially has “child” or “branch”

1 nodes. The child nodes represent any XML elements that occur between the start
2 and end tags of the “parent” node.

3 XML accommodates an infinite number of database schemas. A schema is
4 a set of rules for constraining the structure and articulating the information set of
5 XML documents. A schema describes what data structures, shape, and content of
6 XML documents are valid for a given application. For example, one schema
7 might describe how documents used in an online banking exchange are structured.
8 Other schemas may describe XML documents for email, or XML documents for
9 purchasing blue jeans or music over the Internet.

10 To illustrate a tree structure constructed from XML data, consider an
11 exemplary XML data exchange between different entities, such as client and
12 server computers, in the form of requests and responses. A client might generate a
13 request for information or a request for a certain server action, and a server might
14 generate a response to the client that contains the information or confirms whether
15 the certain action has been performed. The contents of these requests and
16 responses are XML documents. In many cases, the process of generating these
17 XML documents involves building, in memory, a hierarchical tree structure. Once
18 the hierarchical tree structure is built in its entirety, the actual XML document in
19 proper syntactic form can then be assembled. Consider the following exemplary
20 XML code:

21
22 <trans:orders xmlns:person="http://www.schemas.org/people"
23 xmlns:dsig=http://dsig.org
24 xmlns:trans="http://www.schemas.org/transactions">
25 <trans:order>
26 <trans:sold-to>
27 <person:name>
28 <person:last-name>Layman</person:last-name>
29 <person:first-name>Andrew</person:first-name>

```
1      </person:name>
2      </trans:sold-to>
3      <trans:sold-on>1997-03-17</trans:sold-on>
4      <dsig:digital-signature>1234567890</dsig:digital-
5      signature>
6      </trans:order>
7  </trans:orders>
```

8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

This code includes three XML namespace declarations that are each designated with “xmlns”. A “namespace” refers to a dictionary or set of element names defined by the schema. Namespaces ensure that element names do not conflict, and clarify who defined which term. They do not give instructions on how to process the elements. Readers still need to know what the elements mean and decide how to process them. Namespaces simply keep the names straight.

Within an XML document, namespace declarations occur as attributes of start tags. Namespace declarations are of the form “xmlns:[prefix]=[uri]”. A namespace declaration indicates that the XML document contains element names that are defined within a specified namespace or schema. “Prefix” is an arbitrary designation that will be used later in the XML document as an indication that an element name is a member of the namespace declared by universal resource indicator “uri”. The prefix is valid only within the context of the specific XML document. “Uri” is either a path to a document describing a specific namespace or schema or a globally unique identifier of a specific namespace or schema. Uri is valid across all XML documents. Namespace declarations are “inherited”, which means that a namespace declaration applies to the element in which it was declared as well as to all elements contained within that element.

With reference to the above XML code, the namespace declarations include a prefix, e.g. “person”, “dsig”, and “trans” respectively, and the expanded

1 namespace to which each prefix refers, e.g. “<http://www.schemas.org/people>”,
2 “<http://dsig.org>”, and “<http://www.schemas.org/transactions>” respectively. This
3 code tells any reader that if an element name begins with “dsig” its meaning is
4 defined by whoever owns the “<http://www.dsig.org>” namespace. Similarly,
5 elements beginning with the “person” prefix have meanings defined by the
6 “<http://www.schemas.org/people>” namespace and elements beginning with the
7 “trans” prefix have meanings defined by the
8 “<http://www.schemas.org/transactions>” namespace.

9 It is noted that another XML document that incorporated elements from any
10 of the namespaces included in this sample might declare prefixes that are different
11 from those used in this example. As noted earlier, prefixes are arbitrarily defined
12 by the document author and have meaning only within the context of the specific
13 element of the specific document in which they are declared.

14 Fig. 1 shows a hierarchical tree structure 18 that represents the structure of
15 the above XML code. The tree nodes correspond to elements parsed from the
16 XML document. Such a structure is typically constructed in memory, with each
17 node containing all data necessary for the start and end tags of that node. It has
18 been typical in the past to build the entire tree structure before generating the
19 XML document itself.

20 In XML 1.0, data types in the schemas are defined using a set of data type
21 definitions (DTD). XML documents have two kinds of constraints: well-
22 formedness and validity. The “well-formedness” constraints are those imposed
23 by the definition of XML itself (such as the rules for the use of the < and >
24 characters and the rules for proper nesting of elements). The “validity” constraints
25 are constraints on document structure provided by a particular DTD or XML-Data

1 schema. Schema or DTD validation is very useful in the Internet realm, because
2 entities are able to validate whether data structures received from random or
3 anonymous sources are appropriate for a given context. Suppose, for example,
4 that a company receives XML data from some random user. The company does
5 not necessarily trust the data at this point, and hence utilizes a validation process
6 to determine whether the XML data is good or whether it is noise that can be
7 rejected outright or sent to a system administrator for special consideration.

8 Fig. 2 shows current software architecture 20 for processing XML
9 documents. The architecture 20 includes an XML parser 22 that receives and
10 parses XML data. The XML data may arrive in a variety of ways, including as a
11 stream, a URL (universal resource locator), or text. Parsing the XML data results
12 in a list of events. For example, suppose the XML data describes an author and
13 title name for a book, as follows:

14
15 <Book>
16 <author> X </author>
17 <title> Y </title>
18 </Book>

19 The parser 22 parses the XML data and returns the following list:
20
21 1. “Book” element
22 2. BeginChildren
23 3. “Author” element
24 4. BeginChildren
25 5. “X”, text
26 6. EndChildren
27 7. “Title” element
28 8. BeginChildren
29 9. “Y”, text
30 10. EndChildren
31 11. EndChildren

1 As the parser 22 parses the XML data, it calls to one or more node
2 factories. A “node factory” is a callback interface that builds node objects used to
3 construct an in-memory tree representation of the XML document. The node
4 factory may also be used to search the XML document, without building a node
5 object. Custom node factories can be constructed to build different kinds of object
6 hierarchies that reflect the XML document.

7 In architecture 20, there are four node factories, including a namespace
8 node factory 24, a DTD node factory 26, a tree builder node factory 28, and a
9 validation node factory 30. The XML parser 22 calls the namespace node factory
10 24, which outputs a sequence of name tokens. DTD events are passed to the DTD
11 node factory 26 and XML data events are passed to the validation node factory 30.
12 The DTD node factory 26 builds DTD objects 32 from the DTD events. The DTD
13 objects 32 are used in the validation process of validating the XML data. The
14 DTD node factory 26 may also delegate to the tree builder node factory 28, which
15 builds XML DOM (Document Object Model) fragments 34 for pieces of the tree
16 structure, or XML DOM.

17 The validation node factory 30 receives the XML data events from the
18 namespace node factory 24 and uses the DTD objects 32 to evaluate whether the
19 data complies with certain constraints defined by the DTD objects. If the XML
20 data is valid, the tree builder node factory 28 builds a complete XML DOM 36 for
21 the XML data. Some elements of the XML DOM 36 may reference fragments 34.

22 The architecture 20 is configured for DTD-specific considerations. DTD
23 objects have an advantage in that they are known and previously defined. DTD
24 objects also have a drawback, however, in that they are not extensible. Thus, new
25 data type definitions cannot be easily created. Due to this problem, more recent

1 evolutions of XML are beginning to use XML-data schemas as an additional or
2 alternative way to define data types in schemas. XML-data schemas are not
3 restricted like DTD objects, but are extensible and new ones can be created as
4 needed.

5 Thus, there is a desire to adapt the architecture of Fig. 2 to handle the more
6 extensible XML-data schemas in addition to DTD objects. One problem with this
7 adaptation is that the node factory interface provides a sequential ordered stream
8 of XML tokens, whereas the XML-data schemas define items in a way that is
9 order independent. This means that the node factory has to store certain states
10 until it knows it can process those states.

11 Accordingly, there is a need for an improved architecture built around the
12 node factory design, which handles XML-data schemas to build in-memory tree
13 representations and DTD objects for validation purposes. More particularly, the
14 improved architecture should leverage existing components (e.g., DTD validation,
15 namespace node factory, and XML parser) for creating an in-memory
16 representation of the schema and be roughly as fast as the existing architecture.
17 The architecture should also maximize code reuse.

18

19 **SUMMARY**

20 This invention concerns an architecture for processing XML documents
21 and building in-memory tree representations of the XML documents. More
22 particularly, the architecture accommodates XML-data schemas as a way of
23 defining data types within schemas and converts the XML-data schemas to DTD
24 objects that are used to validate the XML data in the XML document.

In the described implementation, the architecture utilizes a node factory design in which an XML parser calls one or more node factories to construct an in-memory tree representation of the XML document. One of the node factories is a schema node factory, which is a thin layer interface that is callable by the parser. The schema node factory translates the parser calls into calls to a schema builder, which is a table drive component that builds DTD objects used in validation of the XML data. If valid, the XML data is used to construct the in-memory tree representation of the XML document.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is diagram of a hierarchical tree structure that represents an XML document.

Fig. 2 is a block diagram of a software architecture for parsing XML documents and creating a tree structure.

Fig. 3 is a block diagram of a client/server system that exchanges XML documents over a network.

Fig. 4 is a block diagram of an XML processing software architecture, including a schema node factory and schema builder, which is implemented in the client and/or server of the system of Fig. 3.

Fig. 5 is a diagram of tables used by the schema builder to convert XML-data schema into DTD objects.

Fig. 6 is a block diagram of a computer system that is suitable for implementing the invention.

Fig. 7 is a flow diagram showing a method for processing an XML document.

1

2 **DETAILED DESCRIPTION**

3 **System**

4 Fig. 3 shows client-server system 40 in which XML documents or data
5 streams are exchanged between a client 42 and a server 44 over a network 46 (e.g.,
6 Internet). The client 42 runs one or more applications that generate requests for
7 XML data and receive responses to those requests in the form of XML documents.
8 One specific example of such client-based applications is Outlook Express from
9 Microsoft Corporation.

10 The client 42 implements an XML processing software architecture 50 that
11 sends, receives, and processes XML documents. The server 44 is equipped with a
12 similar architecture, but will not be described here for brevity purposes. The XML
13 processing unit 50 has a network module 52 (e.g., TCP/IP module) to facilitate
14 connection to the network 46 and a transport object 54 (e.g., a Distributed
15 Authoring and Versioning (WebDAV) transport object) to handle requests and
16 responses over the network. Upon receipt of an XML document, the transport
17 object 54 pushes the XML document into an XML parser 56, which in this
18 example, is a so-called “push-model” parser. The parser 56 parses the XML data
19 stream into individual elements of schema and data. The schema elements
20 establish a schema for an associated application (not shown), while the data
21 element are validated against the schema and if valid, are used to form an in-
22 memory tree representation of the XML document.

23 The parser 56 calls one or more node factories 58 to construct the in-
24 memory tree representation of the XML document. One of the node factories is a
25 “schema node factory” 60, which is a thin layer interface that receives calls from

the parser 54 and delegates the work to another interface known as a “schema builder” 62. The schema builder 62 converts the schema elements contained in the calls to data type definition (DTD) objects. The DTD objects are then used to validate the XML data from the parsed XML document to determine whether the data is valid for the particular schema associated with the application. If valid, the XML data is used to form an in-memory tree representation, which is also known as an XML DOM (Document Object Model) 64.

Architecture

Fig. 4 shows the XML processing software architecture 50 implemented in client 42. The architecture parses an XML document into schema and data, converts the schema to DTD objects, validates the data using the DTD objects, and constructs an in-memory tree representation of the XML document from valid data. The architecture 50 includes the XML parser 56 that receives and parses XML data. The XML data may arrive in a variety of ways, including as a stream, a URL (universal resource locator), or text. The XML data contains one or more schemas and data.

The parser 56 calls a namespace node factory 66 for both the schema(s) and data. A node factory is a callback interface that builds node objects used to construct an in-memory tree representation of the XML document. The general node factory interface (“IXMLNodeFactory”) has several methods, including those listed in Table 1.

Table 1: Node Factory Interface

<u>Method</u>	<u>What it does...</u>
NotifyEvent	This method informs the node factory where an event occurs within the XML document
BeginChildren	This method is called when a node contains children.
EndChildren	This method is called when all sub-elements of the given element are complete.
Error	This method is called when the parser runs into an error in the XML document.
CreateNode	This method is called during parsing for each and every element.

The namespace node factory 66 outputs a sequence of name tokens.

Tokens relating to the schema(s) are passed to the schema node factory 60, whereas tokens pertaining to the data are passed to a validation node factory 68. The schema node factory 60 is a thin node factory interface that receives the schema-related tokens from the namespace node factory 66 and delegates work to a schema builder 62. The schema node factory 60 also passes the tokens onto the tree builder node factory 70 for construction of schema DOM fragments 72 for pieces of the tree representation.

The schema builder 62 is a table-driven schema interface that utilizes static tables 80 to build DTD objects 74. More particularly, the schema builder 62 includes a state machine that is table-driven by values in the static tables 80. The static tables 80 contain information describing the schema schema if you will, including what elements and attributes are valid for XML-Data schemas and the associated functions to call when processing those elements. The static tables 80 are organized in a hierarchy, beginning with a schema root table and cascading

1 down through element and attribute tables and event-to-function mapping tables.
2 The schema builder 62 outputs a list of executable functions that are mapped out
3 by the table hierarchy. The schema builder 62 also has code to construct DTD
4 objects used to test the XML data for validation.

5 The validation node factory 68 receives the data elements from the
6 namespace node factory 66. The validation node factory 68 uses the DTD objects
7 built by the schema builder 62 to evaluate whether the data elements comply
8 with the constraints set forth in the DTD objects. The validation node factory 68
9 passes the valid data elements to the tree builder node factory 70 to build the
10 complete XML DOM 64 for the XML data. Some elements of the XML DOM 64
11 may reference fragments 72.

12 Fig. 5 shows an exemplary table hierarchy of static tables 80 used by the
13 schema builder 62. The table hierarchy includes a schema root table 82 that
14 contains schema information. Root table 82 also contains information on element
15 types and attribute types used in the schema. This information includes a
16 reference (e.g., pointer) to a subordinate table that further describes the types, such
17 as element types table 84 and attribute types table 86.

18 The element types table 84 contains one or more fields listing types of
19 elements permitted in the schema and references to other tables containing
20 information on specific types. The attribute types table 86 contains one or more
21 fields listing types of attributes in the schema, corresponding functions for
22 handling the attributes, and type descriptions.

23 The types tables 84 and 86 may further reference one or more function
24 tables. In Fig. 5, the element types table 84 references (e.g., via a pointer) a
25 function table 88 that contains functions pertaining to the IXMLNodeFactory

1 events described in Table 1. The function table includes functions like
2 ProcessElement, ProcessEndChildren, and ProcessBeginChildren.

3 The schema builder 62 exposes an API (application program interface) to
4 make these and other functions available. In this example, the schema node
5 factory 60 makes calls to the schema builder API, but this API is designed for
6 more general use in other scenarios in addition to the context described here. The
7 schema builder API has several methods, including the methods listed in Table 2.

8

9 **Table 2: Schema Builder Interface**

<u>Method</u>	<u>What it does...</u>
void start();	This method signals that a new schema is being processed and the internal state is reset accordingly.
void finish();	This method signals that an end of the schema has been reached. It checks whether all elements referenced in the schema have been declared.
ProcessElementNode (Node * pNode);	This method is called for each schema element in the document. This occurs when the schema builder looks up the new schema element (e.g., ElementType or AttributeType) in the element types table 84 or attribute types table 86 for the current schema element. If the element is allowed (i.e., listed in the table), the method pushes the current element onto a stack and calls the “init” function defined in the function table 88 for the new schema element. The “init” function initializes one or more data structures for the new schema element and creates empty DTD objects for that element.
ProcessAttributes (IXMLNodeSource * pSource, Node * pNode);	This method is called when all attributes are ready to be processed for the current schema element. This occurs when the schema builder walks through the attributes in the XML DOM 64 for the schema, looks them up in the attribute types table 86 for the current schema element, parses the attribute value according to the

attribute type, and calls the function for handling that attribute.

1 ProcessPCDATA
2 (Node * pNode,
3 PVOID pParent);
4

This method is called for each Text node (including attribute values) in the XML document. This method determines whether it is valid to have text in the current position within the schema document and generates an error if it is not valid.

5 ProcessEndChildren
6 (IXMLNodeSource *
7 pSource,
8 Node * pNode);
9

This method is called when the end of all children for the current schema is reached. This method determines if all content for the current schema element has been correctly provided, along with any final default values. For ElementType elements, the method finishes the content model, which may involve creating a few more DTD content model nodes for describing certain default values.

Exemplary Computer System

Fig. 6 shows a general example of a computer 130 that can be used in accordance with the invention. Various numbers of computers such as that shown can be used in the context of a distributed computing environment. The computer 130 is representative of one implementation of client 42 and/or server 44.

Computer 130 includes one or more processors or processing units 132, a system memory 134, and a bus 136 that couples various system components including the system memory 134 to processors 132. The bus 136 represents one or more of any of several types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, and a processor or local bus using any of a variety of bus architectures. The system memory 134 includes read only memory (ROM) 138 and random access memory (RAM) 140. A basic input/output system (BIOS) 142, containing the basic routines that help to

1 transfer information between elements within computer 130, such as during start-
2 up, is stored in ROM 138.

3 Computer 130 further includes a hard disk drive 144 for reading from and
4 writing to a hard disk (not shown), a magnetic disk drive 146 for reading from and
5 writing to a removable magnetic disk 148, and an optical disk drive 150 for
6 reading from or writing to a removable optical disk 152 such as a CD ROM or
7 other optical media. The hard disk drive 144, magnetic disk drive 146, and optical
8 disk drive 150 are connected to the bus 136 by a SCSI interface 154 or some other
9 appropriate interface. The drives and their associated computer-readable media
10 provide nonvolatile storage of computer-readable instructions, data structures,
11 program modules and other data for computer 130. Although the exemplary
12 environment described herein employs a hard disk, a removable magnetic disk 148
13 and a removable optical disk 152, it should be appreciated by those skilled in the
14 art that other types of computer-readable media which can store data that is
15 accessible by a computer, such as magnetic cassettes, flash memory cards, digital
16 video disks, random access memories (RAMs), read only memories (ROMs), and
17 the like, may also be used in the exemplary operating environment.

18 A number of program modules may be stored on the hard disk 144,
19 magnetic disk 148, optical disk 152, ROM 138, or RAM 140, including an
20 operating system 158, one or more application programs 160, other program
21 modules 162, and program data 164. A user may enter commands and
22 information into computer 130 through input devices such as a keyboard 166 and a
23 pointing device 168. Other input devices (not shown) may include a microphone,
24 joystick, game pad, satellite dish, scanner, or the like. These and other input
25 devices are connected to the processing unit 132 through an interface 170 that is

1 coupled to the bus 136. A monitor 172 or other type of display device is also
2 connected to the bus 136 via an interface, such as a video adapter 174. In addition
3 to the monitor, personal computers typically include other peripheral output
4 devices (not shown) such as speakers and printers.

5 Computer 130 commonly operates in a networked environment using
6 logical connections to one or more remote computers, such as a remote computer
7 176. The remote computer 176 may be another personal computer, a server, a
8 router, a network PC, a peer device or other common network node, and typically
9 includes many or all of the elements described above relative to computer 130,
10 although only a memory storage device 178 has been illustrated. The logical
11 connections include a local area network (LAN) 180 and a wide area network
12 (WAN) 182. Such networking environments are commonplace in offices,
13 enterprise-wide computer networks, intranets, and the Internet.

14 When used in a LAN networking environment, computer 130 is connected
15 to the local network 180 through a network interface or adapter 184. When used
16 in a WAN networking environment, computer 130 typically includes a modem 186
17 or other means for establishing communications over the wide area network 182,
18 such as the Internet. The modem 186, which may be internal or external, is
19 connected to the bus 136 via a serial port interface 156. In a networked
20 environment, program modules depicted relative to the personal computer 130, or
21 portions thereof, may be stored in the remote memory storage device. It will be
22 appreciated that the network connections shown are exemplary and other means of
23 establishing a communications link between the computers may be used.

24 Generally, the data processors of computer 130 are programmed by means
25 of instructions stored at different times in the various computer-readable storage

1 media of the computer. Programs and operating systems are typically distributed,
2 for example, on floppy disks or CD-ROMs. From there, they are installed or
3 loaded into the secondary memory of a computer. At execution, they are loaded at
4 least partially into the computer's primary electronic memory. The invention
5 described herein includes these and other various types of computer-readable
6 storage media when such media contain instructions or programs for implementing
7 the steps described below in conjunction with a microprocessor or other data
8 processor. The invention also includes the computer itself when programmed
9 according to the methods and techniques described below.

10 For purposes of illustration, programs and other executable program
11 components such as the operating system are illustrated herein as discrete blocks.
12 It is recognized that such programs and components reside at various times in
13 different storage components of the computer, and are executed by the data
14 processor(s) of the computer.

15

16 Operation

17 Fig. 7 shows a method for processing an XML document. The steps are
18 performed in software and will be described with additional reference to the
19 architecture 50 of Fig. 4.

20 At step 200, an XML document is received at the parser 56. The XML
21 document consists of XML data that is articulated by and complies with a schema.
22 The parser 56 parses the XML document to produce both data elements and
23 schema elements (step 202).

24 After parsing, the schema elements are converted to DTD objects 74, which
25 are used to validate the XML data (step 204). More particularly, the conversion

1 step involves the parser 56 calling the schema node factory 60 (i.e., CreateNode)
2 to create a node in the tree that represents the schema element (step 206). The
3 schema node factory 60 in turn calls the schema builder interface 62 (e.g., Table 2
4 calls) to process the individual elements and construct data type definition (DTD)
5 objects used in validating the data elements (step 208).

6 To illustrate how the schema node factory 60 and schema builder 62
7 operate, consider the oft-used CreateNode call (Table 1) made from the parser 56
8 to the schema node factory 60. The architecture translates the CreateNode calls
9 into the DTD objects 74. For discussion purposes, suppose the following schema
10 fragment is input:

```
11 <Schema xmlns="urn:schemas-microsoft-com:xml-data">  
12     <ElementType name="foo" model="open">  
13         </ElementType>  
14     </Schema>
```

15 Table 3 lists the series of calls made to the schema node factory 60 and
16 schema builder 62 for this schema fragment.

17
18 **Table 3: Calls Generated In Response for
Parser's CreateNode Call**

<u>SchemaNodeFactory Call</u>	<u>SchemaBuilder Call</u>	<u>Resulting Action</u>
NotifyEvent STARTDOCUMENT	Start	Initialize state and create new empty DTD object.
CreateNode ELEMENT “Schema”	ProcessElementNode	Set current state to point to the Schema element state table.
CreateNode ATTRIBUTE “xmlns”	No operation	

1	CreateNode TEXT “urn:...”	ProcessPCDATA	Attribute text is a no operation.
2	BeginChildren	ProcessAttributes	“xmlns” is a skippable attribute, so this is a no operation.
3			
4	CreateNode ELEMENT “ElementType”	ProcessElementNode	“ElementType” is an allowable child of “Schema” so push “Schema” onto the stack and initialize a new ElementDecl DTD object with default content model EMPTY.
5			
6			
7			
8			
9	CreateNode ATTRIBUTE “name”	No operation	
10			
11	CreateNode TEXT “foo”	ProcessPCDATA	Attribute text is a no operation.
12			
13	CreateNode ATTRIBUTE “model”	No operation	
14			
15	CreateNode TEXT “open”	ProcessPCDATA	Attribute text is a no operation.
16			
17	BeginChildren	ProcessAttributes.	Calls BuildElementName, which initializes name of the ElementDecl DTD object and validates that an ElementDecl named “foo” has not already been declared.
18			
19			
20			
21			Calls BuildeElementModel, which validates the value “open” and sets the model to open.
22			
23	EndChildren “ElementType”	ProcessEndChildren	Calls EndElementType, which determines that there are no child
24			
25			

1		elements in the schema fragment and that the content model is open.
2		Changes the default content model from EMPTY to ANY.
3		Pops current state back to "Schema" State
4		No cleanup function for "Schema" state.
5		Pops "Schema" state back to initial state.
6		
7	EndChildren "Schema"	ProcessEndChildren
8		
9		
10	NotifyEvent ENDDOCUMENT	Finish
11		Checks for referenced but undeclared element types—and there are none.
12		
13		

With reference again to Fig. 7, the data elements from the parsed XML document are evaluated against the DTD objects 74 to determine whether the data elements are valid (step 210). If valid, the tree builder node factory 70 constructs the in-memory tree representation of the XML data (step 212).

Conclusion

Although the invention has been described in language specific to structural features and/or methodological steps, it is to be understood that the invention defined in the appended claims is not necessarily limited to the specific features or steps described. Rather, the specific features and steps are disclosed as preferred forms of implementing the claimed invention.